Powerful, Simple, and Efficient Ensemble Data Assimilation:
Beyond the Kalman Filter

Jeffrey L. Anderson

I. A simple framework for ensemble filtering assimilation

II. Examples of ensemble filtering algorithms
- A. Perturbed Observation Ensemble Kalman Filter
- B. Ensemble Adjustment Kalman Filter
- C. Quadrature Kernel Filters

III. Ensembles for quality control of observations

IV. Examples of abilities of Ensemble Adjustment Kalman Filter
- A. Comparison to adjoint methods in low order model
- B. Application in global PE ‘NWP’ model
- C. Difficult assimilation: global PE model, surface pressure observations only

V. Bias, bias, everywhere...

VI. Future plans

Nonlinear Filtering

Dynamical system governed by (stochastic) DE

$$dx_t / dt = M(x_t, t) + G(x_t, t) w_t \quad (1)$$

Observations at discrete times

$$y_t^o = h_t(x_t, t) + \varepsilon(x_t, t) \quad (2)$$

Observational error is white in time and Gaussian

$$\varepsilon \rightarrow N(0, R) \quad (3)$$

Complete history of observations is

$$Y_\tau = \{ y_l^o ; \; t_l \leq \tau \} \quad (4)$$

Goal: Find probability distribution for state at time t

$$p(x_t | Y_t) \quad (5)$$

Filtering in Joint State/Observation Space

Define joint state/observation vector at time t as $\mathbf{z}_t = \{\mathbf{x}_t, \mathbf{h}_t(\mathbf{x}_t, t)\} = \{x, y\}$

Vector of length n + m: m is size of the observational set $\mathbf{y}^o$ available at time t
n is size of model

State between observation times is obtained from DE (1):

Need to update given new observations at time t:

Define prior joint state: $p(z^p) = p(z_t \mid Y_{t-1})$

updated (posterior) joint state: $p(z^u) = p(z_t \mid Y_t)$

$\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{z}^p \mid \mathbf{y}^o{}_t)$.

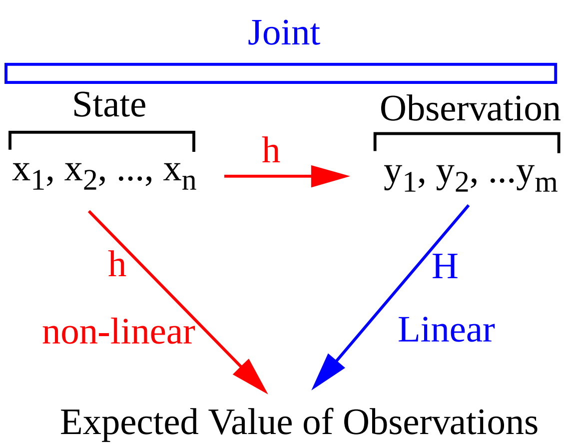
Applying Bayes' rule and fact that observational error is white in time gives:

$$\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{y}^o \mid \mathbf{z}^p)\, \mathbf{p}(\mathbf{z}^p) \,/\, \text{(normalization)} \qquad (6)$$

(drop subscript t from here on)

Can now apply filter directly to arbitrary (non-linear) observation operators

Observations $\mathbf{y}^o$ are related to the joint state variables by simple linear operator $\mathbf{H}$



$$\begin{array}{cc} m \times n & m \times m \\ \end{array}$$
$$H = \begin{vmatrix} 0 & I \end{vmatrix}$$

Sequential Assimilation of Observations

p(a, b | c) = p(a | c) p(b | c) if a and b are independent

Divide observations at given time into s subsets

$y^o = \{y^o_1, y^o_2, \ldots y^o_s\}$

Observational errors for obs. in different subsets independent

First term in numerator equation(6):

$$p(y^o | z^p) = \prod_{i=1}^{s} p(y_i^o | z^p)$$

Can assimilate subsets sequentially in arbitrary order

Gaussian observation errors, subsets have no cross-covariance

Sequential Assimilation of Observations (cont.)

Assume observational system gives errors as Gaussian, mean 0

Observational error covariance matrix is R

Do Singular Value Decomposition (SVD) so that

$$R' = F^T R F$$

where F is unitary.

This is rotation to frame where R is diagonal

Can redefine forward observation operator as $h' = F^T h$

In this frame, all observations have independent error

Only changes definition of observation part of joint space

From now on...
Assimilate single scalar observation without loss of generality

(number of obs. $m = 1$,   size of joint space $k = n + 1$)

Ensemble Filtering

Goal: Obtain as much ‘information’ as possible about p(x)

Assimilation (fusion) process: $\mathbf{p(z}^{u}\mathbf{)} = \mathbf{p(y}^{o}\mathbf{\mid z}^{p}\mathbf{)}\ \mathbf{p(z}^{p}\mathbf{)}$ / (norm)

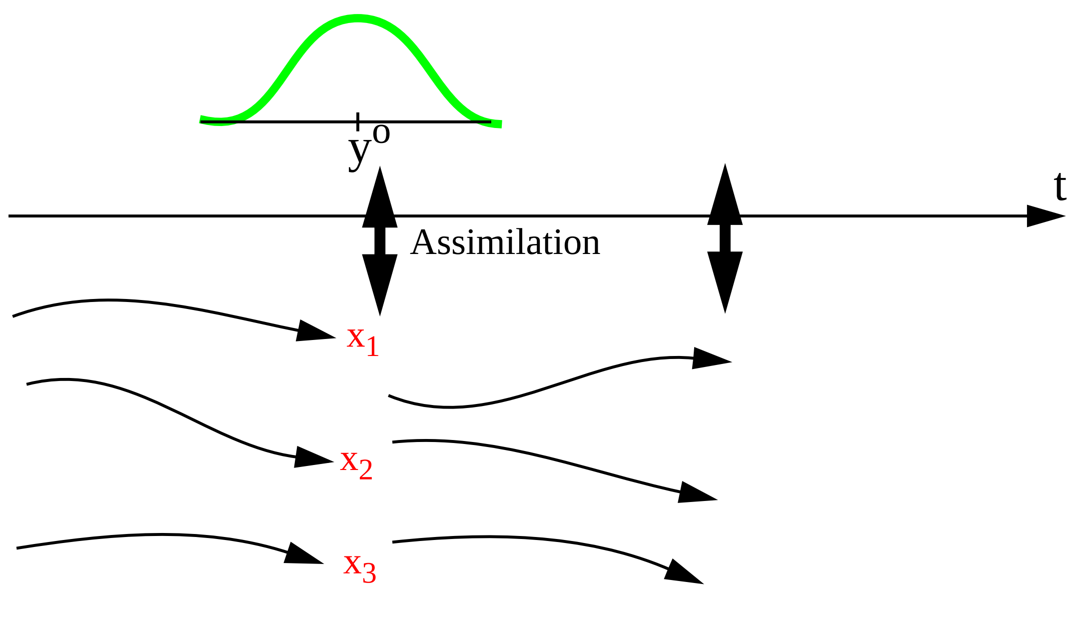
Input:

Observations - unbiased with Gaussian observational error
(Can be easily generalized to sum of Gaussians)

Model: ‘Ensemble’ sample of model prior state distribution

Output:

‘Ensemble sample of model updated (posterior) distribution



Two Step Ensemble Data Assimilation

Step 1: Update marginal distribution for observation variable

$p_y(y^u) = p(y^o \mid y^p)\, p_y(y^p) / (\text{norm})$

Subscript y is marginal distribution on observation variable

Can update y independently of all other variables

Details on y update methods (step 1) later

Result is set of increments for ensemble estimates of y

$y^u_i = y^p_i + \Delta y_i$

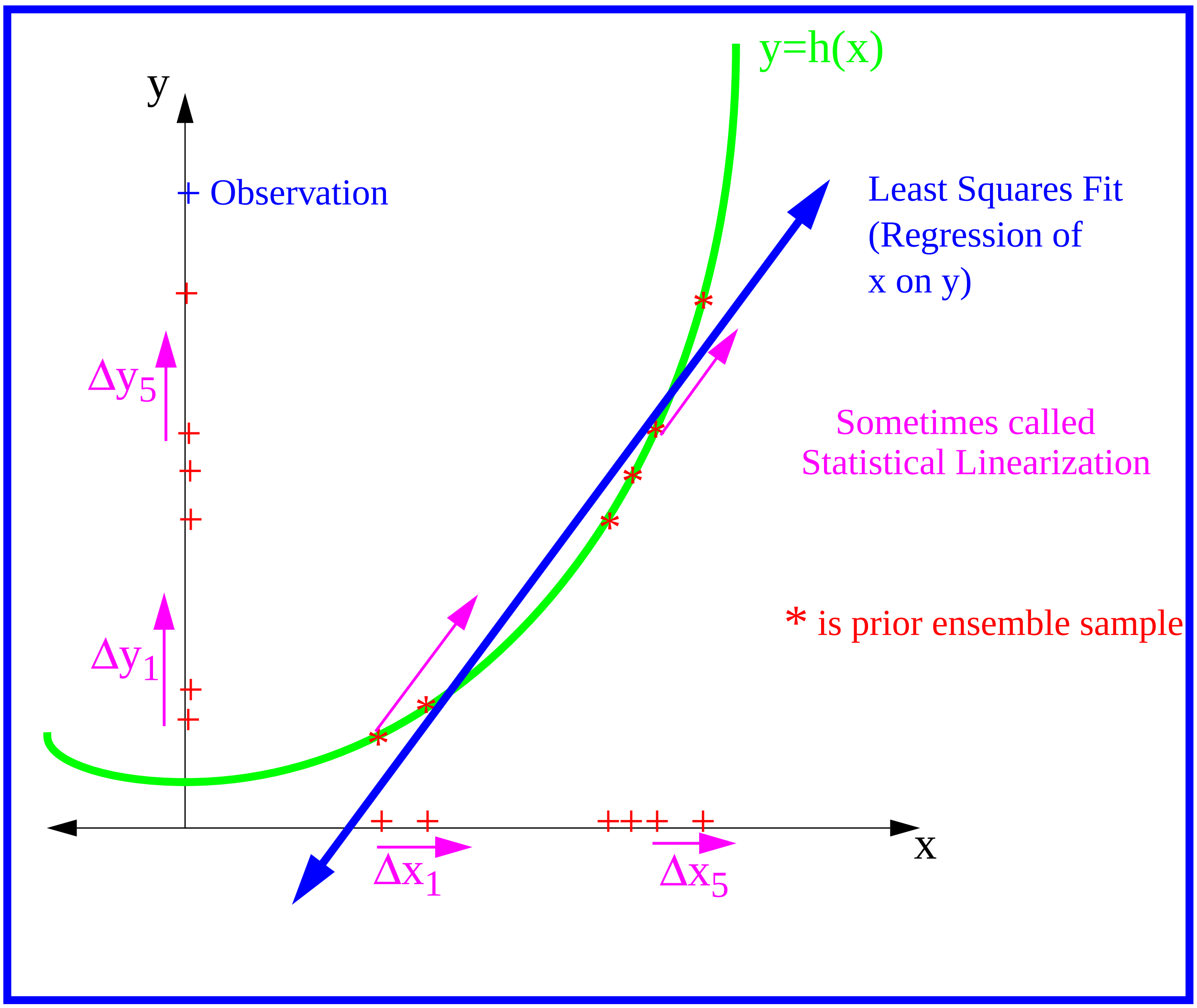
$i = 1, ..., N$ N is ensemble size

Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y, find increments for state variable ensembles

Simple idea: Do linear regression of x^p on y^p

Equivalent to doing: Least squares fit
Assuming Gaussian prior relation
(Doing previously documented ensemble Kalman filters)



Example 1: Observation variable direct function of state variable

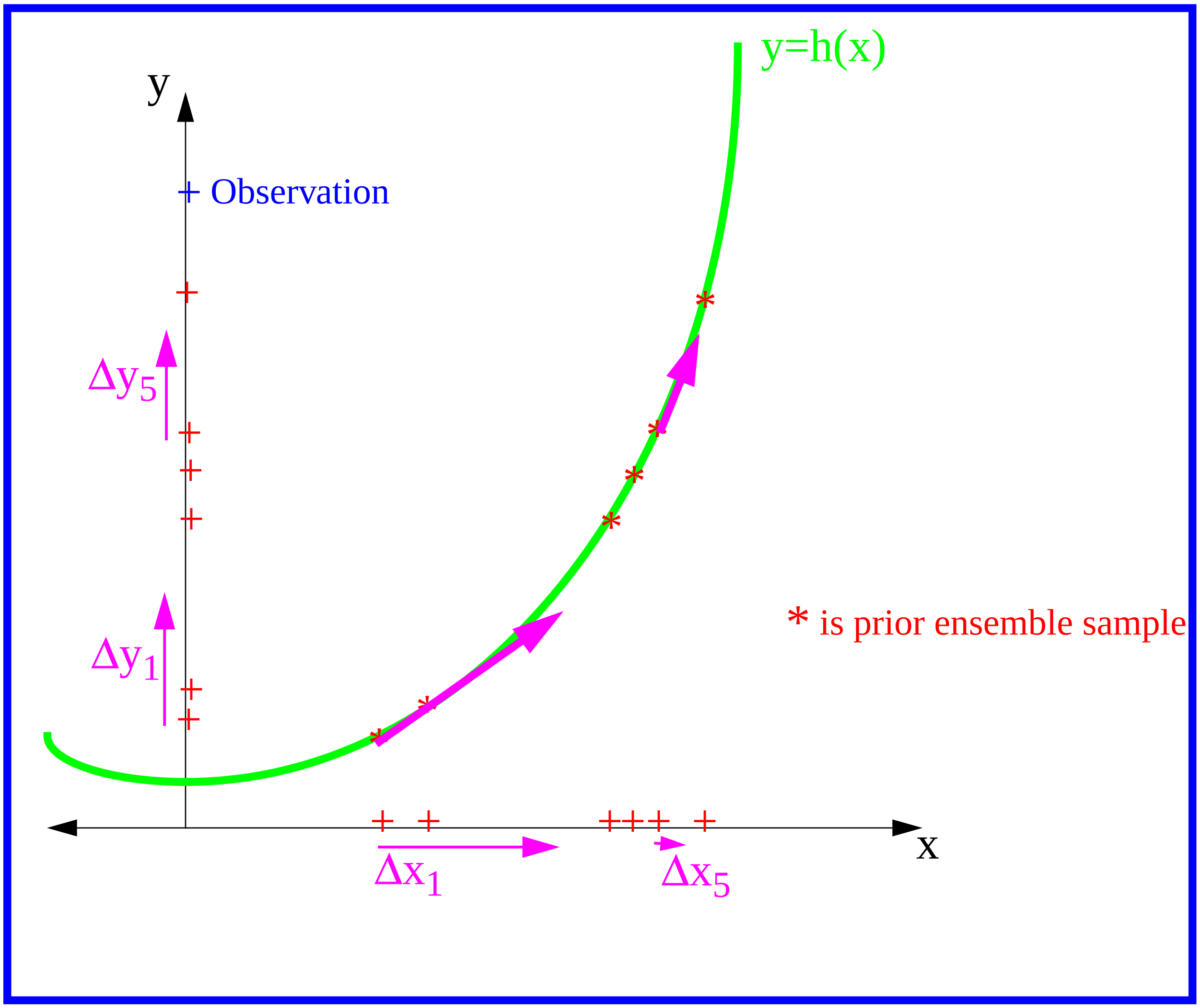
Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y, find increments for state variables

Idea: Do linear regression of x^p on y^p

Could also do local linearizations:

(Related to Gaussian Kernel approximation)



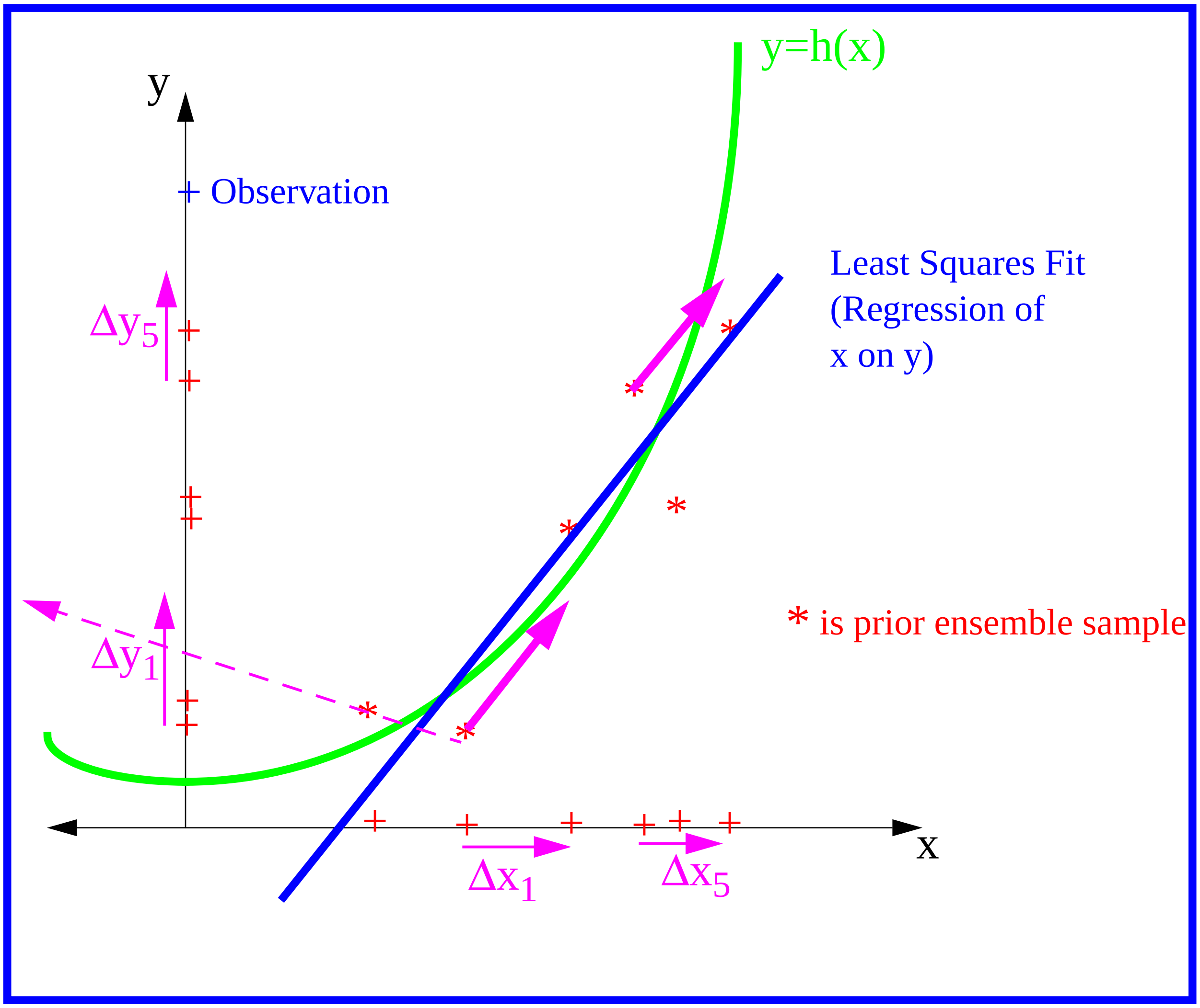
Works well in cases where state and obs are functionally related

Two Step Ensemble Assimilation (cont.)

Step 2: Given increments for y, find increments for state variables

More challenging when obs and state are not functionally related

Example: $y = h(x_2)$, x and x_2 strongly correlated



Large sample size needed to ‘remove’ noise

Trade-offs with local linearization (dotted magenta)

Details of Step 1: Updating Observation Variable Ensemble

Scalar Problem: Wide variety of options available and affordable

Begin with two previously documented methods:

1. Perturbed Observation Ensemble Kalman Filter

2. Ensemble Adjustment Kalman Filter

Both make use of following (key to Kalman filter...)

Given prior ensemble with sample mean $\overline{z}^p$ and covariance Σ^p

Observation y^o with observational error variance matrix R

Note: Product of Gaussians is Gaussian

$$\Sigma^u = \left\{ \left(\Sigma^p\right)^{-1} + H^T R^{-1} H \right\}^{-1} \quad (9)$$

and mean:

$$\overline{z^u} = \Sigma^u \left\{ \left(\Sigma^p\right)^{-1} \overline{z^p} + H^T R^{-1} y^o \right\} \quad (10)$$

Details of Step 1: Perturbed Obs. Ensemble Kalman Filter

1. Compute prior sample variance and mean, Σ^p and $\bar{z}^p$

2. Apply (9) once to compute updated covariance, Σ^u

3. Create an N-member random sample of observation distribution by adding samples of obs. error to y^o

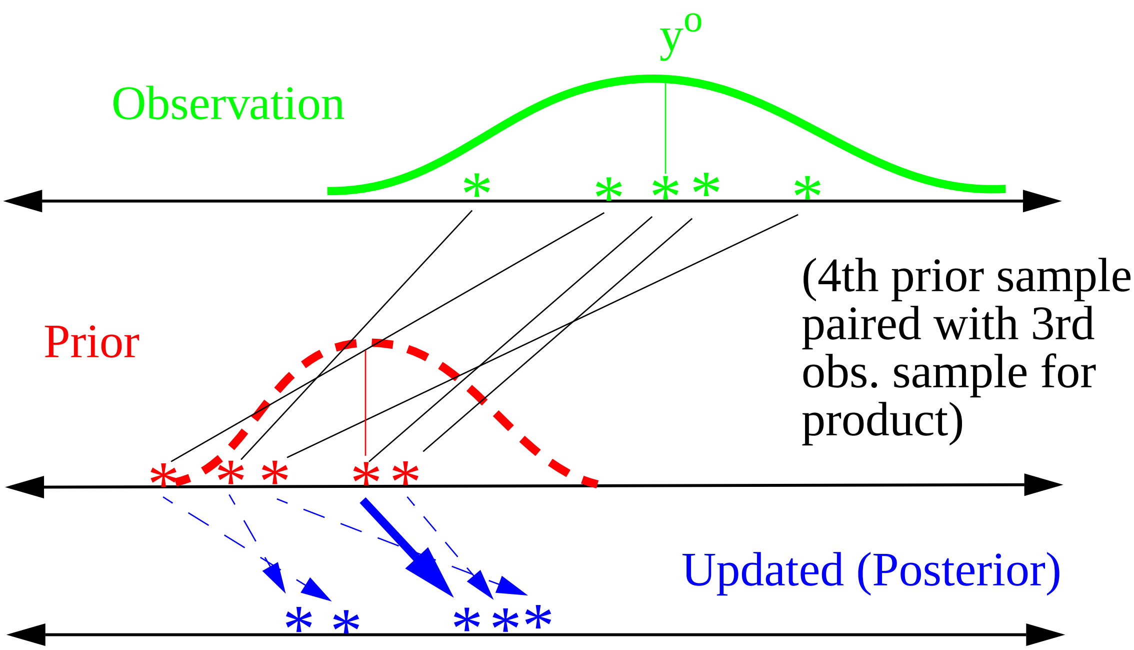
4. Apply (10) N times to compute updated ensemble members

   Replace $\bar{z}^p$ with value from prior ensemble, $y^p{}_i$

   Replace y^o with value from random sample, $y^o{}_i$

   Updated ensemble value is $y^u{}_i$ (= $\bar{z}^u$ from 10)

NOTE: When combined with linear regression for step 2, this gives identical results to EnKF's described in literature!



Details of Step 1: Ensemble Adjustment Kalman Filter

1. Compute prior sample variance and mean, Σ^p and $\bar{z}^p$

2. Apply (9) once to compute updated covariance, Σ^u

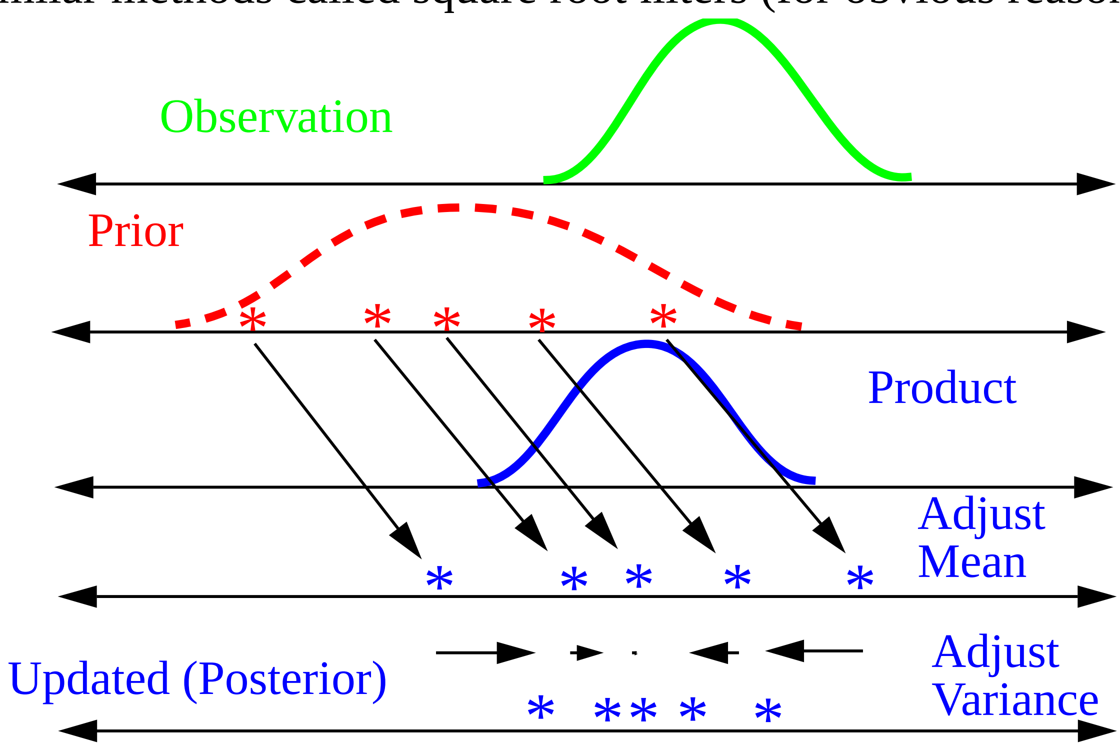
3. Apply (10) to compute updated mean, $\bar{z}^u$

4. Adjust prior ensemble of y so that mean and variance are exactly $\bar{z}^u$ and Σ^u

$$y_i^u = \left(y_i^p - \bar{y}^p\right)\sqrt{\sigma^u / \sigma^p} + \bar{y}^u, \quad i = 1, \ldots N$$

σ is variance of y

Similar methods called square root filters (for obvious reasons)



Details of Step 1: Quadrature kernel filter

1. Compute prior sample variance Σ^p

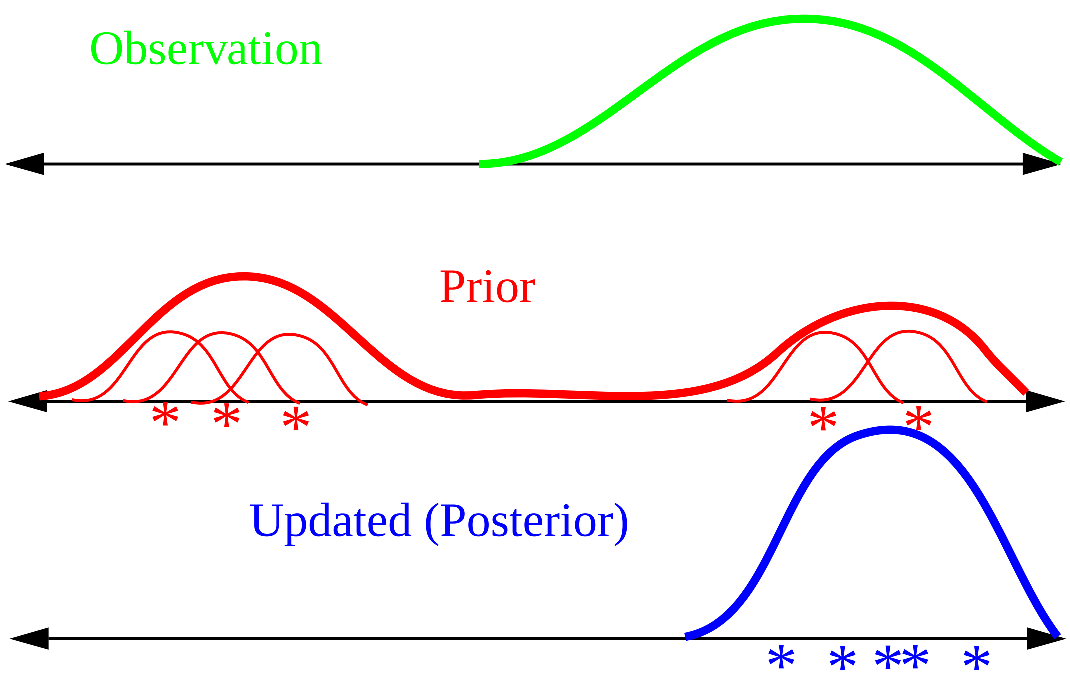
2. Use a Gaussian (or other) kernel approximation to get
   continuous approximation to $p(y^p)$

3. Use quadrature to take product in (6) directly
   Can do individual Gaussians kernels if Gaussian
   $\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{y}^o \mid \mathbf{z}^p)\, \mathbf{p}(\mathbf{z}^p)$ / (normalization)

4. Create an N-member random sample of $p(z^u)$

5. An interesting variant uses boxcar kernels

6. Only useful for non-Gaussian structure in prior;
   very powerful for Lorenz-63 model



Limiting the Impact of Observations

Only letting each observation impact subset of state vars in
step 2 regression is often advantageous

1. Reduces computational cost

2. Avoids problems due to spurious remote correlations
(Especially problematic for small ensembles, large state)
(Hamill et al., 2001)

3. Avoids singularity problems from small sample sizes
(gives updates in more than n -1 directions)
This is crucial in realistic models with bias

Extremely straightforward in two step filter context
- Just multiply covariance in regression by a distance-dependent factor
- Close observations have full impact
- Remote observations have no impact
- Could rephrase this argument in terms of expected inter-relatedness as opposed to distance

Quality Control of Observations

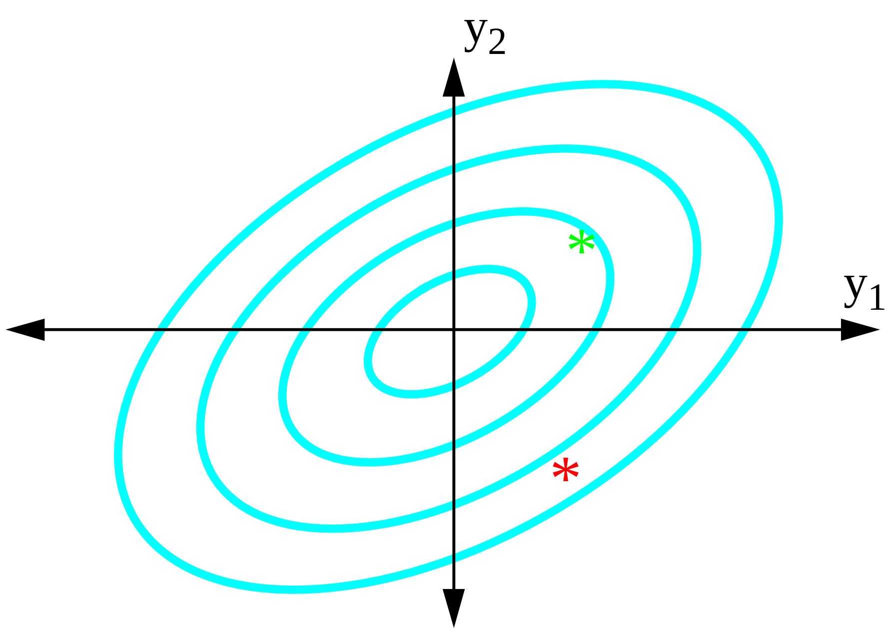
Methods to exclude erroneous observations

1. Discard impossible values (negative R.H.)

2. Discard values greatly outside climatological range

3. Discard values that are more than α prior ensemble sample standard deviations away from prior ensemble mean

4. ‘Buddy’ checks for pairs of observations: just apply chi-square test using prior ensemble covariance and label pair as inconsistent if threshold value exceeded



5. Could also apply chi-square to larger groups of obs.

Comparison to Adjoint: Computational Cost and Implementation

1. Model integrations
   a. Filter requires N forward integrations of model; O(10) sufficient?
   b. Adjoint requires K*L forward and backward integrations
      K - number of observation intervals over which optimization is performed
      L - average number of iterations of minimization solver
      K*L at least O(10) for any envisioned application

2. Assimilation algorithm cost
   a. Filter: $O(\alpha Nnm)$:N is ensemble size, n is model size, m is number of obs
      α related to what fraction of state variables are impacted by given ob.
      In certain scenarios this may reduce order of cost
   b. Adjoint: O(nm) in best of all possible cases
      Relation of constant factors not clear, depends on ensemble size

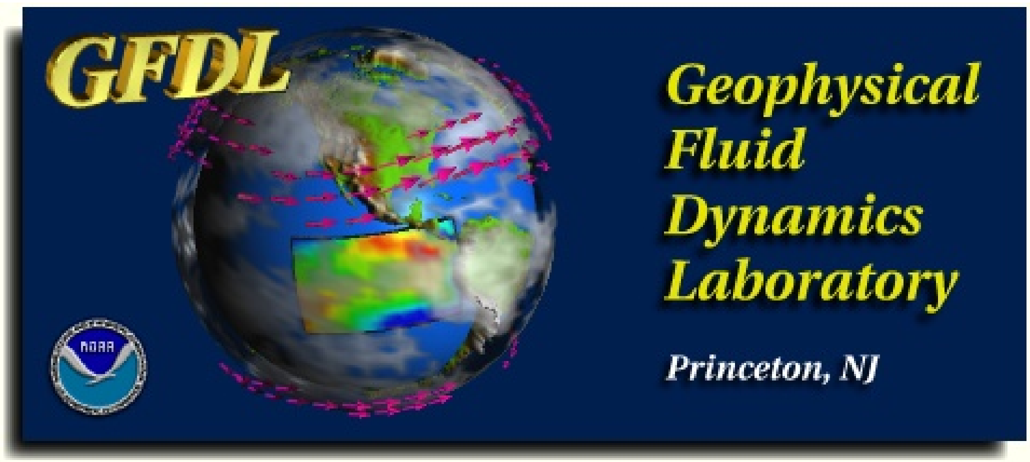
3. Ease of implementation
   a. Filter-needs no model specific software
   b. Adjoint-requires exact adjoint model plus linear tangent (can be a pain)

4. Ability to deal with model imperfection
   a. Filter-single parameter adjusts relative confidence of prior / obs.
      Easy model parameter assimilation may help introduce noise
   b. Adjoint-additional difficulties to implement model as weak constraint
      Addition of model noise very tricky without ensemble information

5. Information produced
   a. Filter gives information on mean and distribution
   b. Adjoint gives information on ‘mean’

6. Adjoint can be extended in a number of ways...

Ensemble Filter Results in Global PE Models



PE models components of GFDL’s Flexible Modeling System

B-grid dynamical core by Bruce Wyman

B-grid model incorporated for filter by Shaoqing Zhang

Global GCM results by Shaoqing Zhang

Global B-grid General Circulation Model

Dynamical Core (Distant cousin of NCEP ETA model):

- Global B-grid
- σ/p hybrid vertical coordinate
- Prognostic equations:
  - Momentum (u & v)
  - Temperature (T)
  - Specific Humidity (q)
  - Surface Pressure (p_s)
- Resolution: 90 longitudes, 60 latitudes, 18 levels (N30L18)
  - (n = 486,000)

Parameterizations:

- Long- and short-wave radiation (Fels-Schwarzkopf)
- Moist convective adjustment and large-scale condensation
- Mellor-Yamada 2.5 vertical turbulence
- Stern-Pierrehumbert gravity wave drag
- Simple land surface with bucket hydrology

Observational network:

- 600 column observations distributed randomly on sphere
- Quantities observed on models vertical levels
- Synthetic observations taken from perfect model control
- Observations every 12 hours
- Obs. error standard deviation 1C, 1m/s, 100hPa

Ongoing Work

1. Explore methods for dealing with model systematic error in mean and covariance

2. Explore implications for initialization in intermediate models

3. Explore targeted observation methodology (Shree Khare)

4. Implement ensemble adjustment filter in MOM ocean model for SI prediction (Shaoqing Zhang)

5. A variety of fun OSSEs? More on use of surface obs.

Future Plans:

Explore applications in atmospheric ensemble prediction systems

Examine evolving stochastic parameterizations, tuning parameterizations with ensemble assimilation

Incorporate into experimental data assimilation test-bed to be developed at NCAR in collaboration with OAR